

# Projeto de Pesquisa para Iniciação Científica

## Estruturação de um modelo evolutivo baseado no indivíduo

Candidata: Irina Lerner

Orientador: Carlos Eduardo Ferreira<sup>1</sup>

Co-Orientadora: Flávia Maria Darcie Marquitti<sup>2</sup>

<sup>1</sup> Instituto de Matemática e Estatística – Usp

<sup>2</sup> Instituto de Física ‘Gleb Wataghin’ – Unicamp

### Resumo

Modelos baseados em indivíduo (IBM) são usados atualmente para o estudo de evolução, por permitirem modelar tanto aspectos no nível individual, populacional e de comunidades. Por estas características, este tipo de modelagem permite estabelecer uma ligação entre forças micro-evolutivas, como por exemplo mutações genéticas e fluxo genético, e os padrões observados no nível macroevolutivo, como por exemplo através da estrutura de árvores filogenéticas. Baseado na teoria Neutra da Biodiversidade, o modelo proposto por Aguiar *et al.* 2009 pode ser usado para estudar como uma população inicialmente homogênea se diversifica em várias espécies devido às mutações, à reprodução sexuada e à estruturação dos indivíduos no espaço. Neste modelo, cada indivíduo é caracterizado por um genoma, definido por uma *string* binária. Os indivíduos são separados em grupos chamados de espécies dependendo do fluxo genético entre os indivíduos. Como é um modelo evolutivo, a cada geração a identificação das espécies deve ser feita. O custo computacional da reprodução sexuada e da busca pela espécie é elevado quando consideramos muitos indivíduos e/ou genomas grandes, já que essas funções dependem da comparação entre os genomas dos indivíduos. Neste trabalho, implementaremos tal modelo, estruturando-o de forma que seja possível o uso de genomas grandes e muitos indivíduos. Em um próximo passo, buscaremos entender como a variação de alguns parâmetros do modelo, tal como o parâmetro relacionado ao fluxo genético, influenciam na formação de espécies e quais os padrões que podem ser observados em árvores filogenéticas.

**Palavras-chave:** modelagem em grafos, modelo baseado no indivíduo, especiação

# 1 Introdução

Entender como a diversificação de uma população de indivíduos ocorre e leva à separação destes indivíduos em espécies é uma das perguntas essenciais em biologia (Gavrilets, 2014). Estudos teóricos e empíricos sobre o surgimento e a importância da diversidade estão presentes em diferentes níveis de organização. Do ponto de vista de um indivíduo, estudos evolutivos são usados para compreender como se dá a evolução de diversas estratégias entre células, como por exemplo estratégias cooperativas e egoístas, que culminam na formação de indivíduos multicelulares (Dingli et al., 2009). No nível populacional, a caracterização das frequências de diversos alelos em sub-populações nos ajudam a entender como são mantidos polimorfismos e como podemos caracterizar uma espécie (Charlesworth et al., 1997; Guillot et al., 2005). No nível de espécies, o interesse evolutivo pode ser por exemplo, entender como a invasão por uma certa espécie resulta em uma radiação (adaptativa ou não) dando origem a novas diversas espécies (Taylor et al., 2006; Rundell and Price, 2009). Recentes avanços computacionais possibilitaram a elaboração de modelos teóricos mais complexos, permitindo unir áreas muito distantes, tal como estabelecer a relação entre processos micro evolutivos e os padrões macroevolutivos associados a estes processos (Barraclough and Vogler, 2000; Pigot et al., 2010; Gavrilets, 2014; Gascuel et al., 2015; Costa et al., 2019). É nesta ligação micro-macroevolutivo que se insere este projeto de pesquisa.

O processo de especiação, em geral, é um processo de milhões de anos e, por perdemos algumas informações que não são registradas em fósseis ou são difíceis de se interpretar a partir de dados moleculares, os processos evolutivos e as assinaturas deixadas por estes processos em árvores filogenéticas podem ser estudados através de modelos (Gavrilets, 2003) matemáticos, estatísticos ou baseados de simulações. Entre os modelos de simulações está a modelagem baseada no indivíduo ou no agente (IBM ou ABM — Individual/Agent Based Model) (Gavrilets et al., 2000; Dieckmann and Doebeli, 1999; Pigot et al., 2010; Gascuel et al., 2015). Neste tipo de modelo evolutivo, os próprios indivíduos ou até agentes de maior escala taxonômica (como populações ou espécies) são modelados individualmente (Grimm et al., 2006). A partir da interação entre tais agentes, processos emergentes, como por exemplo a especiação, começam a ocorrer. O modelo proposto por Aguiar e colaboradores (2009) é um modelo evolutivo baseado no indivíduo, e neste trabalho, ele será implementado.

No modelo,  $N$  indivíduos possuem genomas binários de mesmo tamanho ( $B$ ), e estão dispostos espacialmente em uma região limitada. Eles podem se deslocar de acordo com um raio, um dos parâmetros do modelo. Neste deslocamento, encontram possíveis parceiros (outros indivíduos) com quem cruzam se a distância entre seus genomas é menor que um parâmetro  $G$ , deixando um descendente, caso cruzem. O genoma deste novo indivíduo é uma combinação dos genomas parentais. Pode haver ainda a probabilidade, em cada *locus* do genoma, de mutação. Em seguida, os indivíduos parentais morrem. Este modelo é dito neutro, e usa uma abordagem derivada da Teoria Neutra da Biodiversidade (Hubbell, 2001). Isto é, não considera efeitos adaptativos ou de seleção natural, ou em outras palavras, características não são favorecidas ou desfavorecidas na sobrevivência ou reprodução individual.

A modificação dos parâmetros do modelo, tal como tamanho do genoma, raio de busca por parceiros, distância reprodutiva  $G$  e tamanho do espaço, permite simular diferentes situações de processos evolutivos. Em especial, a variação no raio de busca por parceiros muda o processo geográfico de evolução a ser estudado. Os processos de especiação geográficos estão relacionados com o estabelecimento e interrupção de fluxo genético e têm sido utilizados para explicar como se deu a especiação de diferentes táxons. O processo de especiação alopátrico, em que se estabelece uma barreira geográfica que impede o fluxo genético entre grupos de indivíduos que ficam separados pela barreira (Yamaguchi and Iwasa, 2013), é uma das teorias mais aceitas. Estes grupos isolados acumulam diferenças, como por exemplo por meio da deriva genética (Lande, 1981; Uyeda et al., 2009), diferenciação no uso de recursos (Grant and Grant, 2008; Losos and Ricklefs, 2009) ou adaptações locais (Barton and Polechová, 2005; Schluter, 2009), que culminam na formação de espécies. Estudos recentes e modelos de simulação de evolução tem buscado evidências e também compreender os mecanismos pelos quais espécies se formam quando o fluxo genético não é interrompido, como ocorrem nos processos de especiação geográfica simpátrico (Dieckmann and Doebeli, 1999; Coyne and Orr, 2004) e parápátrico (Gavrilets et al., 2000; Coyne and Orr, 2004). Ambos ocorrem quando não há separação dos indivíduos através de barreiras geográficas e o fluxo gênico entre esses novos grupos não é interrompido. Na simpatria não há restrição alguma do fluxo genético, e na parapatria, as barreiras são parciais.

A cada geração é calculado o número de espécies existentes. Para isso, é preciso uma definição de espécie, conceito com muitas nuances. Uma das mais usadas é a proposta por Mayr e conhecida como conceito biológico de espécie que define estas como agrupamentos de populações naturais intercruzantes, reprodutivamente isolados de outros grupos com as mesmas características (Mayr, 1942). Neste trabalho, usaremos o conceito de fluxo gênico entre indivíduos para definir as espécies (Petit and Excoffier, 2009). Isto é, buscamos grupos que estão isolados reprodutivamente do restante dos indivíduos. Desta forma, estabelecemos como espécie o grupo de indivíduos que estão a uma certa distância genética máxima entre si, ou seja, são possíveis parceiros sexuais, ou que estabelecem o fluxo gênico via possíveis parceiros em comum. Esta definição exige que se verifique todos os indivíduos que pertencem a um grupo onde existe fluxo genético.

Este modelo é destacado por sua eficiência computacional, e por fazer uma ponte entre a microevolução, representada por mutações pontuais nos genomas dos indivíduos e a reprodução sexual entre eles, e a macroevolução, representada pelas relações de parentesco entre as espécies (árvores filogenéticas) e distribuição espacial de cada espécie (De Aguiar et al., 2009; Costa et al., 2019). Apesar disso, nas implementações anteriores, a busca de espécies, e a busca por parceiros sexuais são ainda muito custosas e limitantes no tempo de execução de algumas simulações, em especial aquelas cujos parâmetros de tamanho populacional e genoma são muito grandes. A razão disso é que encontrar espécies depende de uma comparação entre os  $N$  genomas de tamanho  $B$ , que pode ser feito em tempo  $O(N^2 B^2)$ , e encontrar parceiros sexuais pode ter complexidade  $O(N * B)$ . Este limitante é ainda mais custoso quando se estuda a evolução em processos simpátricos. Para a simulação

de um processo simpátrico, por causa do alto fluxo genético entre todos os indivíduos, necessitamos estabelecer um genoma muito grande. Portanto, o estudo da importância da especiação simpátrica e comparações deste processo com a parapatria e alopatria passam pela necessidade de se solucionar esse problema, ou seja, implementar o modelo de forma que a função que busca parceiros e que define são as espécies entre todos os indivíduos a cada geração sejam otimizadas.

Em uma primeira etapa, buscaremos implementar o modelo de Aguiar e colaboradores estabelecendo uma estrutura de dados para definir cada indivíduo da população. Esta estruturação permitirá que estabeleçamos um grafo (ou uma rede de interações), definido como  $\mathcal{G}(V, A)$ , onde  $V$  é um conjunto de vértices, que aqui representam os indivíduos da população, e  $A$  é o conjunto de pares de vértices não ordenados (ou as arestas), entre todos os indivíduos estabelecendo conexões entre os indivíduos que são geneticamente compatíveis entre si (isto é, possíveis parceiros sexuais). A partir dessa definição em grafo, buscar parceiros e definir as espécies a cada geração são informações contidas no grafo. A tarefa de buscar a espécie no grafo, em especial, resume-se à tarefa de encontrar o número de subgrafos maximais conexos (ou componentes) na estrutura total (Bondy et al., 1976), o que possivelmente diminui o custo computacional anteriormente mencionado. Esta busca no grafo de fluxo genético pode ser otimizada através de algoritmos de busca em largura ou de busca em profundidade em grafos (Hopcroft and Tarjan, 1973). A diminuição do custo computacional permitirá simulações de processos de especiação simpátrica e, em uma próxima etapa, entender os sinais deixados pela evolução em simpatria e até entender como ocorre a evolução da característica que define o raio de busca por parceiros sexuais em um modelo neutro de evolução.

## 2 Justificativa

Diversos estudos teóricos vêm sendo desenvolvidos com o objetivo de se entender como processos evolutivos, tais como a geografia, a adaptação e as alterações ambientais, moldam o surgimento (e desaparecimento) de espécies. O modelo que propomos estudar e implementar preenche uma lacuna entre os estudos evolutivos, pois permite modelar processos ao nível dos indivíduos. Estes processos deixam assinaturas de escala macroevolutiva, tais como número de espécies, distribuição das abundâncias das espécies e estruturas das árvores filogenéticas.

Um dos entraves do modelo atualmente é estudar processos geográficos de simpatria e poder comparar os padrões deste processo com outros processos geográficos de evolução. Isto ocorre pois as simulações em simpatria exigem um tamanho de genoma muito grande. Além disso, os componentes estocásticos do modelo exigem que as análises de cada conjunto de parâmetros sejam feitas com muitas réplicas para se observar as tendências médias, aumentando ainda mais o tempo de execução total necessário para se entender os padrões do processo simpátrico de especiação. Neste sentido, o desenvolvimento deste projeto contribuirá para que o modelo evolutivo possa ser otimizado em tempos de execução e posteriormente possa ser utilizado na comparação com outros processos evolutivos.

Outro ponto importante é que o estudo desse problema permitirá à aluna solicitante desenvolver

conhecimentos nas áreas de teoria dos grafos e biologia evolutiva. A aluna solicitante é estudante do curso de bacharelado em Ciências Moleculares pela Universidade de São Paulo, que possui caráter interdisciplinar tal como o projeto proposto e os orientadores associados. Assim sendo, o projeto preenche os requisitos para uma iniciação científica interdisciplinar a ser executado por uma pesquisadora em formação.

### 3 Objetivos

- Implementar o modelo usando grafos: Estabelecer um grafo de fluxo genético, ligando indivíduos que são possíveis parceiros sexuais.
- Estudar a complexidade dos algoritmos propostos ao serem implementados, comparar os tempos de execução no modelo evolutivo e propor novas modificações que possam diminuir a complexidade do programa em funções ou instantes de tempo específicos.
- Complementar a formação da aluna na interface entre Ciência da Computação e Biologia.

### 4 Cronograma

Apresentamos abaixo as atividades que serão realizadas ao longo do período de Iniciação Científica, considerando o projeto com duração de 12 meses.

1. Estudo de conceitos básicos e aspectos de implementação de algoritmos em grafos e IBM's
2. Estudo teórico de modelagem de evolução
3. Implementação do programa em linguagem C
4. Implementação de diferentes funções buscando otimizar a busca de parceiros sexuais e a definições das espécies
5. Análise dos resultados obtidos
6. Implementação de bibliotecas em R e Python para o IBM com a finalidade de disponibilizar o modelo para usuários
7. Relatório final

O cronograma previsto para a execução das atividades acima listadas estão alocadas entre o 1º mês e o 12º mês do período de Iniciação Científica de acordo com a Tabela 1.

Tabela 1: Cronograma												
Atividade	Mês											
	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11°	12°
1	•	•										
2	•	•	•									
3			•	•	•							
4					•	•	•	•				
5									•	•		
6										•	•	
7											•	•

## 5 Materiais e Métodos

O modelo utilizado para a simulação de evolução é baseado no indivíduo, como descrito na Subseção 5.1. Nas Subseções seguintes, descrevemos uma estruturação de dados para o modelo (Subseção 5.2), possíveis formas de se efetuar a otimização de funções que buscam parceiros sexuais para reprodução e a busca para a definição das espécies (Subseção 5.3), além de um aprimoramento na computação das diferenças genéticas entre os indivíduos (Subseção 5.4).

### 5.1 Modelo baseado no indivíduo

O modelo baseado no indivíduo que será utilizado é o proposto por Aguiar e colaboradores (2009). Este é um modelo considerado neutro na evolução já que não há diferenças entre os indivíduos que definam as suas chances de se reproduzir e deixar descendente, isto é, não há uma aptidão diferencial entre os indivíduos. Todos os indivíduos (sendo um total de  $N$  indivíduos) são definidos por um genoma bialélico –  $\gamma$ , representado por uma *string* de tamanho  $B$  posições ou *loci* ( $\{\gamma_1, \dots, \gamma_i, \dots, \gamma_B\}$ ), onde  $\gamma_i = \{0, 1\}$ . Cada posição ou *locus*  $i$  desse genoma se refere a um gene ou característica do indivíduo. Todo indivíduo tem chance de se reproduzir se: (i) não morrer com uma probabilidade  $D$  e (ii) encontrar um parceiro sexual em uma vizinhança de raio  $R$  em torno dele.

A possibilidade de um indivíduo ser um parceiro sexual de um certo indivíduo focal é definido através da distância genética. A distância genética ( $d^{m,n}$ ) é definida pela distância de Hamming entre os genomas do indivíduo focal ( $\gamma^m$ ) e do seu parceiro ( $\gamma^n$ ), isto é, a soma de todas as diferenças entre os genomas:

$$d^{m,n} = \frac{1}{2} \sum_{k=1}^B |\gamma_k^m - \gamma_k^n| \quad (1)$$

Não há nenhuma preferência para que o indivíduo parceiro tenha a menor distância genética. Esta distância somente deve estar abaixo de um certo valor de parâmetro  $G$ , que define a compatibilidade genética. Ao se reproduzir com um parceiro sexual, o filho do pai focal herdará o *locus*  $i$  do genoma do pai focal ou do parceiro do pai focal com igual probabilidade. Cada *locus* do genoma pode ainda sofrer uma mutação a uma taxa  $\mu$  após a herança parental. Para manter a densidade populacional

constante e uniformemente distribuída ao longo das gerações permite-se que um vizinho aleatório dentro do raio  $R$  se reproduza no lugar de um indivíduo focal que não se reproduziu. Após todos os indivíduos terem tido sua chance de reprodução, computamos uma geração. Não existe sobreposição de gerações, isto é, a geração atual (parental) morre e a geração seguinte é composta somente por seus filhos. Inicialmente (na primeira geração), os indivíduos são genomicamente idênticos e uniformemente distribuídos em um reticulado de largura  $L$  com condições periódicas de contorno. A cada geração, verificamos se houve especiação e quais indivíduos pertencem a quais espécies.

### Busca e definição das espécies

Utilizando uma representação em grafos, podemos definir um grafo de fluxo genético  $\mathcal{G}(V, A)$ , onde os vértices ( $V$ ) são os indivíduos. As arestas ( $A$ ) representam a ligação entre possíveis parceiros sexuais, isto é, ligamos os vértices que representam indivíduos que estão a uma distância genética menor ou igual do que  $G$  ( $d^{m,n} \leq G$ ). Desta forma, as espécies, de acordo com a definição adotada (fluxo genético), são os componentes desse grafo, isto é, são os subgrafos maximais conexos do grafo total. Um subgrafo maximal conexo (componente) é composto por um subconjunto de vértices do grafo total que estão conectados entre si por pelo menos um caminho (Bondy et al., 1976). Portanto existe fluxo genético (ou ao menos um caminho) entre quaisquer indivíduos (vértices) de uma mesma espécie (componente). Veja a ilustração na Figura 1 de um pequeno número de indivíduos divididos em três espécies utilizando este conceito de fluxo genético. Nesta situação, os indivíduos que possuem uma aresta os conectando são possíveis parceiros sexuais neste instante de tempo.

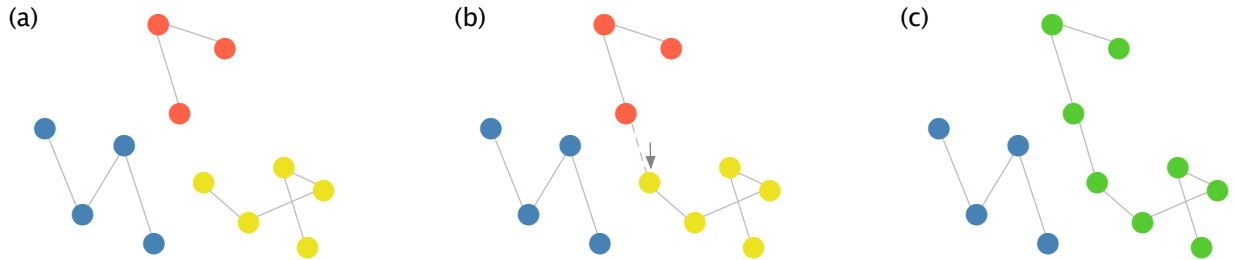


Figura 1: Na figura acima, Em (a) observamos indivíduos (vértices) divididos em três espécies (componentes). Note que nem todos os indivíduos da espécie azul, por exemplo estão conectados por aresta (fluxo genético). No entanto, há pelo menos um caminho que une qualquer indivíduo azul a outro. Em (b), um indivíduo da espécie amarela (indicado pela seta) sofre uma mutação e estabelece fluxo gênico com um indivíduo da espécie vermelha. Em (c) concluímos as novas espécies depois do evento ocorrido em (b), em que o fluxo genético estabelecido por um par de indivíduos faz com que a definição de espécies mude.

## 5.2 Estruturação de dados

Para implementar o programa, primeiramente é necessário definir uma estrutura de dados que modele os indivíduos. Os atributos de cada indivíduo são:

- Nome do indivíduo
- genoma
- posição (x,y) no espaço
- espécie

Para utilizarmos de forma ótima esta estrutura, podemos guardar os indivíduos que são vizinhos espaciais em forma de lista como um quinto atributo do indivíduo. Podemos também guardar a lista de parceiros sexuais como um sexto atributo de cada indivíduo. Desta forma, a estrutura do indivíduo fica:

- Nome do indivíduo
- genoma
- posição (x,y) no espaço
- espécie
- vizinhos espaciais
- parceiros sexuais

### 5.3 Funções de busca das espécies e de busca de parceiros sexuais

Como a reprodução só é possível entre indivíduos que estão em um raio espacial  $R$  de um certo indivíduo focal escolhido e a uma distância genética menor que  $G$ , o atributo **vizinhos espaciais** pode ser utilizado no momento da reprodução para escolher um indivíduo que esteja na vizinhança espacial. Após esta escolha, a verificação se o indivíduo escolhido é um possível parceiro sexual pode ser feita pelo atributo **parceiros sexuais**. Esta estruturação economiza na função da reprodução sexuada que atualmente tem um custo da ordem de  $NB$ , já que cada um dos  $N$  indivíduos precisa ter seu genoma de tamanho  $B$  comparado ao do parceiro sorteado (pela Eq. 1) para se verificar a compatibilidade genética.

As listas de **parceiros sexuais** de todos os indivíduos permitem estabelecer o grafo de fluxo genético na população como um todo, como definido na Subseção 5.3. Com o grafo definido, a definição de espécie de cada indivíduo é feita por uma busca em largura ou pela busca em profundidade dos subgrafos maximais conexos (Hopcroft and Tarjan, 1973). A busca em largura ou em profundidade por componentes de um grafo tem o custo da ordem de  $V + A$  (Hopcroft and Tarjan, 1973), onde esperamos que haja uma economia computacional considerável em relação à atual implementação do modelo.

### 5.4 Computação das distâncias genéticas

A computação das distâncias genéticas entre pares de indivíduos é da ordem de  $B$ . Uma forma de se melhorar a eficiência desta função é manter uma matriz com as distâncias genéticas entre



todos os pares de indivíduos. Além disso, é possível também se economizar tempo computacional ao manter somente a informação dos *loci* gênicos do filho que foram alterados durante o processo de reprodução sexual e mutação em relação ao genoma do pai focal. Portanto as comparações genômicas para se alterar a matriz de distâncias genéticas por este método serão feitas somente nas posições (*loci*) que foram alteradas após reprodução e mutação. Ainda que manter a matriz de distâncias genéticas tenha complexidade  $\mathcal{O}(N^2B)$ , o fato de se comparar somente as alterações pode ser uma importante economia, em especial em casos que o genoma ( $B$ ) é grande. Para tanto, podemos criar um sétimo atributo na estrutura de cada indivíduo, chamado **modificações do genoma**, em que anotamos as posições (*loci*) onde ocorreram modificações no genoma em relação ao genoma do pai focal.

## 6 Forma de análise dos dados

Estudaremos a complexidade computacional das implementações propostas para determinar a eficiência assintótica de cada função. Como este é um modelo evolutivo, é possível que os piores casos não ocorram em cenários práticos. Por exemplo, um caso bastante complexo seria o caso em que todos os indivíduos são geneticamente muito diferentes uns dos outros a ponto de serem cada um de uma espécie distinta. No entanto, dadas as condições do modelo evolutivo, esta situação nunca ocorrerá (ou ocorrerá com baixíssima probabilidade).

Levando em consideração a complexidade estudada teoricamente e a compreensão de detalhes do algoritmo, como por exemplo entender os momentos (gerações) e as partes de cada uma das funções que são computacionalmente mais complexas, poderemos propor novas otimizações para este modelo. Uma vez que os algoritmos estiverem implementados, faremos testes de desempenho, verificando o tempo de execução dependendo dos parâmetros de entrada do modelo (tais como  $N$ ,  $B$  e  $R$ ).

## 7 Elaboração de bibliotecas para R e Python

A linguagem C, aqui escolhida para a implementação do modelo baseado no indivíduo, permite que as manipulações de memória e estruturas de dados sejam personalizadas e possam ser otimizadas da maneira que convém para este problema científico específico. Por outro lado, linguagens interpretadas, tais como R e Python, são atualmente bastante utilizadas em áreas de interface, tais como as áreas entre biologia, matemática e computação. Desta forma, propomos elaborar bibliotecas para as linguagens R e Python para que o modelo que implementaremos em C possa ser utilizado de forma compilada dentro destas linguagens. Destacamos que esta é uma recomendação dada pelo R-project (<https://www.r-project.org/about.html>), em que os idealizadores aconselham a implementação de programas computacionalmente intensivos em linguagens C, C++ ou FORTRAN. Com esta disponibilização, pretendemos tornar o código mais acessível para usuários das áreas de Biologia e Matemática acostumados com outras linguagens e interfaces.

## Referências

- Barracough, T. G. and A. P. Vogler, 2000. Detecting the geographical pattern of speciation from species-level phylogenies. *The American Naturalist* 155:419–434.
- Barton, N. H. and J. Polechová, 2005. The limitations of adaptive dynamics as a model of evolution. *Journal of Evolutionary Biology* 18:1186–1190.
- Bondy, J. A., U. S. R. Murty, et al., 1976. Graph theory with applications, vol. 290. Citeseer.
- Charlesworth, B., M. Nordborg, and D. Charlesworth, 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research* 70:155–174.
- Costa, C. L., P. Lemos-Costa, F. M. Marquitti, L. D. Fernandes, M. F. Ramos, D. M. Schneider, A. B. Martins, and M. A. de Aguiar, 2019. Signatures of microevolutionary processes in phylogenetic patterns. *Systematic Biology* 68:131–144.
- Coyne, J. A. and A. Orr, 2004. Speciation. Sinauer Associates, Sunderland, MA.
- De Aguiar, M. A. M., M. Baranger, E. Baptestini, L. Kaufman, and Y. Bar-Yam, 2009. Global patterns of speciation and diversity. *Nature* 460:384.
- Dieckmann, U. and M. Doebeli, 1999. On the origin of species by sympatric speciation. *Nature* 400:354.
- Dingli, D., F. Chalub, F. Santos, S. Van Segbroeck, and J. Pacheco, 2009. Cancer phenotype as the outcome of an evolutionary game between normal and malignant cells. *British Journal of Cancer* 101:1130–1136.
- Gascuel, F., R. Ferrière, R. Aguilée, and A. Lambert, 2015. How ecology and landscape dynamics shape phylogenetic trees. *Systematic biology* 64:590–607.
- Gavrilets, S., 2003. Perspective: models of speciation: what have we learned in 40 years? *Evolution* 57:2197–2215.
- , 2014. Models of speciation: where are we now? *Journal of heredity* 105:743–755.
- Gavrilets, S., H. Li, and M. D. Vose, 2000. Patterns of parapatric speciation. *Evolution* 54:1126–1134.
- Grant, P. R. and B. R. Grant, 2008. How and why species multiply: the radiation of Darwin’s finches. Princeton University Press, Princeton, NJ.
- Grimm, V., U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S. K. Heinz, G. Huse, et al., 2006. A standard protocol for describing individual-based and agent-based models. *Ecological modelling* 198:115–126.
- Guillot, G., A. Estoup, F. Mortier, and J. F. Cosson, 2005. A spatial statistical model for landscape genetics. *Genetics* 170:1261–1280.
- Hopcroft, J. and R. Tarjan, 1973. Algorithm 447: efficient algorithms for graph manipulation. *Communications of the ACM* 16:372–378.
- Hubbell, S. P., 2001. The unified neutral theory of biodiversity and biogeography (MPB-32). Princeton University Press, Princeton, NJ.
- Lande, R., 1981. Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences* 78:3721–3725.
- Losos, J. B. and R. E. Ricklefs, 2009. Adaptation and diversification on islands. *Nature* 457:830.
- Mayr, E., 1942. Systematics and the origin of species. columbia univ. press, new york. Systematics and the origin of species. Columbia Univ. Press, New York. .

- Petit, R. J. and L. Excoffier, 2009. Gene flow and species delimitation. *Trends in Ecology & evolution* 24:386–393.
- Pigot, A. L., A. B. Phillimore, I. P. Owens, and C. D. L. Orme, 2010. The shape and temporal dynamics of phylogenetic trees arising from geographic speciation. *Systematic biology* 59:660–673.
- Rundell, R. J. and T. D. Price, 2009. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution* 24:394–399.
- Schluter, D., 2009. Evidence for ecological speciation and its alternative. *Science* 323:737–741.
- Taylor, E., J. Boughman, M. Groenenboom, M. Sniatynski, D. Schluter, and J. Gow, 2006. Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology* 15:343–355.
- Uyeda, J. C., S. J. Arnold, P. A. Hohenlohe, and L. S. Mead, 2009. Drift promotes speciation by sexual selection. *Evolution: International Journal of Organic Evolution* 63:583–594.
- Yamaguchi, R. and Y. Iwasa, 2013. First passage time to allopatric speciation. *Interface focus* 3:20130026.